



УДК 004.9: 378.14

DOI: <http://dx.doi.org/10.21686/1818-4243-2024-6-13-21>

Т.М. Шамсутдинова

Башкирский государственный аграрный университет, Уфа, Россия

# Цифровой след как источник больших данных (Big Data) в образовании

**Цель данного исследования** – рассмотреть перспективы и проблемы применения больших данных в образовании.

**Материалы и методы.** Методами исследования выступали анализ, систематизация и структурирование информации в области применения больших данных в образовании, а также моделирование и программная реализация тестовой модели обработки больших данных с использованием фреймворка Apache Spark.

**Результаты.** В статье рассмотрены ключевые аспекты использования больших данных в образовании, в частности, их источники в виде цифрового следа обучения, методы анализа и направления применения больших данных.

При этом были выделены следующие источники больших данных в образовании: электронная образовательная среда и электронная библиотека вуза; мобильные приложения для обучения; сайт вуза; социальные сети и форумы; данные обратной связи, обращений и опросов; персональные данные, включая психометрические характеристики обучающихся; данные научных smart лабораторий; данные видеонаблюдений и систем контроля управления доступом; данные о карьерном пути и успешности выпускников.

Направления применения больших данных в образовании включают в себя следующие моменты: персонализация электронного обучения, выдача персонализированных рекомендаций; аналитика данных; оценка и обратная связь; прогнозирование успеха студентов; мониторинг качества образования; создание модели обучаемого; разработка учебных планов на основе запросов работодателей; разработка новых образовательных программ; появление новых моделей обучения; совершенствование процессов управления вузом; совершенствование работы приемной компа-

нии; модернизация программно-технических средств обучения; оптимизация педагогического состава.

В качестве проблем применения больших данных в образовании рассматриваются проблемы защиты личных (персональных) данных, необходимость в новых методологиях и технологиях анализа больших данных, потребность в существенной модернизации имеющихся в системе образования технических средств, необходимость в квалифицированных кадрах. В статье также приведен тестовый пример анализа log-файла (журнала событий) электронного курса с помощью технологий обработки больших данных Spark SQL, показывающий потенциальную возможность и практическую применимость технологий обработки больших данных к задачам анализа цифрового следа обучения.

**Заключение.** Большие данные в образовании способны предоставить уникальные возможности для анализа и оптимизации учебного процесса, помогая выявить тенденции, прогнозировать успехи студентов и адаптировать образовательные программы к индивидуальным потребностям учащихся. Но нельзя также забывать, что использование больших данных в сфере образования также несет с собой определенные риски и вызовы, связанные с этическими аспектами, защитой личных данных и необходимостью кадровой модернизации сложившейся системы образования. Для успешной интеграции аналитики данных в образовательную практику необходимо развивать не только технические ресурсы, но и уровень цифровой безопасности и этики в использовании персональных данных.

**Ключевые слова:** цифровой след, большие данные, образование, электронное обучение, анализ данных, Spark SQL.

Tatiana M. Shamsutdinova

Bashkir State Agrarian University, Ufa, Russia

# Digital Footprint as a Source of Big Data in Education

**The purpose of this study** is to consider the prospects and problems of using big data in education.

**Materials and methods.** The research methods include analysis, systematization and structuring of information in the field of big data application in education, as well as modeling and software implementation of a test model for processing big data by using the Apache Spark framework.

**Results.** The article considers the key aspects of using big data in education. In particular, the sources of big data in the form of a digital footprint of learning, methods of analysis and areas of application of big data are examined.

At the same time, the following sources of big data in education were identified: electronic educational environment and electronic library of the university; mobile applications for learning; university website; social networks and forums; feedback data, requests and surveys; personal data, including psychometric characteristics of students; data from scientific smart laboratories; data from video surveillance and access control systems; data on the career path and success of graduates.

The use of big data in education includes the following points: personalization of e-learning, issuance of personalized recommendations; data analytics; assessment and feedback; predicting student success; monitoring of education quality; creation of a learner model; development of curricula based on employer requests; development of new educational programs; emergence of new learning models;

improvement of university management processes; improvement of the work of the admission office; modernization of software and hardware teaching aids; optimization of the teaching staff.

The problems of protection of personal data (personal information), the need for new methodologies and technologies for analyzing big data, the need for significant modernization of the technical means available in the education system, the need for qualified personnel are considered as problems of using big data in education. The paper also provides a test example of analyzing a log file (event log) of the e-course using Spark SQL big data processing technologies. The example shows the potential and practical applicability of big data processing technologies to the tasks of analyzing the digital footprint of learning.

**Conclusion.** Big data in education can provide unique opportunities for analyzing and optimizing the educational process, helping to identify trends, predict student success and adapt educational programs to the individual needs of students. However, the use of big data in the educational sphere also implies certain risks and challenges related to ethical aspects, protection of personal data and the need for personnel modernization of the existing education system. For the successful integration of data analytics into educational practice, it is necessary to develop not only technical resources, but also the level of digital security and ethics in the use of personal data.

**Keywords:** digital footprint, big data, education, e-learning, data analysis, Spark SQL.

## Введение

С развитием информационных систем и платформ онлайн-обучения возникает новая эпоха в образовательном пространстве, основанная на принципах персонализации и аналитики. Одним из наиболее значимых аспектов этой трансформации является использование больших данных, которые представляют собой обширные массивы информации, генерируемые в процессе обучения и взаимодействия различных участников образовательного процесса.

Большие данные в образовании способны предоставить уникальные возможности для анализа и оптимизации учебного процесса, помогая выявить тенденции, прогнозировать успехи студентов и адаптировать образовательные программы к индивидуальным потребностям учащихся. Они включают в себя не только традиционные данные об успеваемости, но и информацию о поведении студентов на различных платформах, их предпочтениях, взаимодействиях с учебными материалами и др.

Различные аспекты применения больших данных в образовании России и стран СНГ рассматриваются в работах таких современных авторов как Ю.К. Александрова, Е.Ю. Алексейчева, К.А. Баранников, З.Б. Батчаева, Е.В. Бебенина, О.М. Елкин, В.Л. Гойко, Л.А. Зейналова, А.Б. Кондратенко, С.М. Лесин, Г.А. Мамедова, Р.Т. Меликова, И.Ш. Мухаметзянов, М.А. Назаренко, Е.В. Нехорошева, Е.Ю. Огурцова, Е.М. Сергейчик, И.Б. Тесленко, О.А. Фиофанова, Е.В. Ширинкина и многих других.

Подробный обзор зарубежных исследований в области применения больших данных в образовании приводится в работе M.I. Baig, L. Shuib, E. Yadegaridehkordi [1]. При этом говорится, что текущие

мировые исследования охватывают четыре основные темы: поведение и результативность учащихся, моделирование и хранение образовательных данных, совершенствование образовательной системы, а также интеграцию больших данных в учебную программу.

Но, тем не менее, нельзя считать проблему применения больших данных в образовании полностью исследованной и решенной. Анализ больших данных на основе интеллектуальных информационных технологий представляет собой сложную задачу, включающую целый ряд требований к уровню алгоритмических, программных и технических средств.

Несмотря на множество преимуществ, применение больших данных в образовании также сопряжено с определенными вызовами, включая вопросы этики, защиты персональных данных и необходимости обучения преподавателей использованию аналитических инструментов.

Цель данного исследования – рассмотреть перспективы и проблемы применения больших данных в образовании. В статье рассмотрены ключевые аспекты использования больших данных в образовании, их источники в виде цифрового следа обучения, методы анализа больших данных, направления применения больших данных в образовании, а также приведен тестовый пример анализа log-файла (журнала событий) электронного курса с помощью технологий обработки больших данных Spark SQL.

### 1. Большие данные в образовании: возможности и вызовы

Очевидно, что в последние годы образование претерпевает значительные изменения под влиянием цифровых технологий. Использование больших данных стало важным ин-

струментом, способствующим трансформации учебного процесса. Огромные массивы информации, генерируемые учащимися, преподавателями и образовательными платформами, открывают новые горизонты для анализа и оптимизации технологий образования.

Что же такое большие данные?

Считается, что начало терминологии в области больших данных (Big Data) было положено в 2008 году Клиффордом Линчем (Clifford Lynch) в его известной статье «Big Data: How do your data grow?», опубликованной в журнале «Nature» [2]. Под большими данными при этом будем понимать объемные и сложные массивы информации, которые невозможно эффективно анализировать традиционными методами обработки. Эти данные находятся в концепции 3V (*volume, velocity, variety* – англ., в переводе: объем, скорость, разнообразие), могут включать различные типы информации, такие как числовые значения, текстовые сообщения, видео, аудиофайлы и др.

Источники больших данных в образовании разнообразны и могут включать в себя как традиционные, так и инновационные технологии обучения и взаимодействия обучающихся. Все эти источники данных базируются на так называемом цифровом следе обучения и создают огромный потенциал для анализа и улучшения образовательного процесса. Цифровой след обучения при этом представляет собой разнообразные данные, которые формируются в процессе образовательной деятельности и взаимодействия учащихся с различными образовательными платформами и ресурсами.

В широком понимании этого термина, цифровой след – «это электронная форма представления данных о результатах учебной, профессиональной и социальной деятельности че-

**Источники больших данных в образовании**  
**Sources of big data in education**

Источник	Виды больших данных
Электронная образовательная среда вуза	log-файлы (журналы событий) отчетов о деятельности в электронных курсах: данные о поведении пользователей, их успеваемости, запись времени входа и выхода, данные о действиях учащихся, включая посещаемость, выполнение заданий, взаимодействие с учебными материалами, прохождение тестов. Результаты промежуточной и итоговой аттестации, в том числе итоговое оценивание, агрегирующее информацию о достижениях студентов, а также информацию о модулях, требующих дополнительного/повторного изучения.
Электронная библиотека вуза	Данные о посещении электронной библиотеки, скачивании учебных материалов и методических указаний. Данные об эффективности учебных материалов, о частоте работы с библиотекой. Данные о взаимодействии обучающихся с электронными учебными материалами (выбор электронных учебников, показатели скорости просмотра страниц, отказов, возвраты к страницам, количество просмотров страниц и т.д.).
Мобильные приложения для обучения	Данные о взаимодействии обучающихся со специальным образовательным контентом.
Сайт вуза	Данные о посещении сайта вуза, фиксация участников, совершивших какое-либо действие в пространстве общих коммуникаций.
Социальные сети и форумы	Данные об активности в социальных сетях. Данные из образовательных групп, обсуждений и комментариев на платформах, специализированных форумах для последующего анализа вовлеченности и интересов учащихся.
Обратная связь, обращения и опросы	Опросы и анкетирование как форма оценки мнений всех участников образовательного процесса (студентов, преподавателей, администрации, представителей работодателей и т.д.) о процессе обучения и его качестве. Данные об обращениях граждан, поступающих в органы образования по информационным системам.
Персональные данные, включая психометрические данные	Анализ познавательных и эмоциональных аспектов обучения, таких как мотивация, стресс и эмоциональное состояние учащихся, формирующее представление об индивидуальных потребностях студента, о его взаимодействиях и предпочтениях.
Данные научных smart лабораторий	Данные виртуальных научных smart лабораторий (включая разнообразные датчики, дроны).
Данные видеонаблюдений	Показатель вовлеченности в учебный процесс, показывающий заинтересованность и активность студентов в ходе учебных занятий – какие используют ресурсы, насколько активно участвуют в обсуждениях и групповой работе.
Системы контроля управления доступом	Автоматическая фиксация посещаемости и пропусков занятий.
Данные о карьерном пути и успешности выпускников	Информация о трудоустройстве и карьере студентов-выпускников как источник данных для оценки практико-ориентированности образовательных программ.

ловека, характеризующей уровень его профессиональной компетентности с точки зрения траектории личностного и профессионального развития» [3].

Е.Ю. Огурцова, Р.Н. Фадеев в работе [4] выделяют следующие виды больших данных в образовании:

- данные, отражающие взаимодействие студентов с электронными системами обучения;

- персональные (личные) данные обучающихся;

- данные об эффективности используемых учебных материалов;

- административные (общесистемные) данные.

В работе [5] говорится о получении больших данных от массовых открытых онлайн-курсов и образовательных платформ. Там же отмечается возможность формирования больших данных в виде сведений о персональных особенностях личности обучающихся, их деятельности в социальных сетях, в виде сведений о статусе и квалификации выпускников и др.

В статье [6] говорится о формировании массивов больших данных по результатам мониторинга обращений граждан в системе образования.

Можем заключить, что источниками больших данных в образовании могут выступать самые разнообразные виды и технологии образовательной и научной деятельности (табл. 1).

## 2. Направления применения больших данных в образовании

Применение больших данных открывает широкие перспективы в системе образования.

В частности, Г.А. Мамедова, Л.А. Зейналова, Р.Т. Меликова в работе [7] отмечают, что данные технологии помогут «разрабатывать индивидуальные траектории обуче-

ния и адаптировать процесс обучения для удовлетворения потребностей каждого обучающегося». Анализ таких данных позволит улучшить модель обучающегося, получить подробную информацию о характеристиках обучающегося или его состояниях (например, знаниях, мотивации), поможет изучить, какие конкретные факторы

влияют на повышение усвоения учебного материала.

Ю.В. Ермачкова, М.А. Ливенец, И.В. Селиверстова в своей работе [8] отмечают такой важный аспект применения Big Data, как использование больших данных при прогнозировании запроса рынка труда к системе образования.

**Направления применения больших данных в образовании**  
**Areas of application of big data in education**

Направление	Характеристики направления
Персонализация электронного обучения, выдача персонализированных рекомендаций	Информация о предпочтениях и моделях обучения студентов может использоваться для создания адаптивных образовательных технологий, которые подстраиваются под индивидуальные потребности каждого учащегося. На основе истории обучения система может предлагать персонализированные траектории и ресурсы для дальнейшего развития, что позволяет каждому учащемуся продвигаться в своем индивидуальном темпе.
Аналитика данных	Сопоставление и анализ больших объемов данных позволяют выявлять тенденции, предпочтения и закономерности в обучении, что помогает в разработке более эффективных образовательных программ. В качестве методов анализа могут выступать: построение моделей учащихся, визуализация данных, классификация, кластеризация, регрессия, анализ текстов, ассоциативные правила, анализ социальных сетей, выявление выбросов, прогнозирование, интеллектуальный анализ процессов и взаимосвязей, статистические методы, интеллектуальный анализ данных, поиск паттернов (шаблонов) обучения и др.
Оценка и обратная связь	Автоматизированные системы могут собирать и анализировать информацию о прогрессе в обучении, данные о результатах выполнения заданий, данные о плохо усваиваемых темах, требующих повышенного внимания.
Прогнозирование успеха студентов	На основе анализа данных можно предсказать успеваемость учащихся и определить тех, кто нуждается в дополнительной информационной и методической поддержке.
Мониторинг качества образования	Системы сбора и анализа данных могут помочь в оценке качества образовательных программ, выявлении сильных и слабых сторон, а также в принятии обоснованных решений на уровне учебных заведений и образовательных политик. Оценка качества учебы в разрезе по курсам, включая качество выпускных квалификационных работ. Оценка структуры контента курса и его эффективности в процессе обучения. Степень удовлетворенности участников процесса образования ходом занятий и своими результатами.
Создание модели обучаемого	Включает анализ поведения и уровня обученности учащихся (прогресс и профиль), а также анализ уровня сформированности их профессиональных компетенций. Позволит получить подробную информацию о характеристиках обучаемого (например, его знаниях, мотивации), изучить факторы, влияющие на повышение/понижение усвоения учебного материала.
Разработка учебных планов на основе запросов работодателей	Позволит проектировать актуальные для рынка труда образовательные программы, используя для этого опрос специалистов-эйчаров, анализ массива вакансий на карьерных порталах, анализ требований работодателей, а также прогнозирование трендов запроса рынка труда к системе образования.
Разработка новых образовательных программ	Модернизация и оптимизация образовательной траектории в трендах вызовов времени. Повышение эффективности спроса образовательных программ и определение эффективности подхода к обучению. Оптимизация образовательных программ, позволяющая адаптировать программы и курсы в соответствии с реальными потребностями студентов.
Появление новых моделей обучения	Разработка специфических инструментов анализа данных для обеспечения образовательных запросов, анализ вовлеченности студентов в учебный процесс. Собранные данные помогут выявить тенденции активности студентов и определить, какие методы обучения оказывают наибольшее влияние на их успехи. Своевременно полученная информация об обучающихся позволит реализовать модели гибкого обучения и оперативно реагировать на любые изменения в учебном процессе, вносить требуемые изменения в календарно-тематический план и учебный контент.
Совершенствование процессов управления вузом	Совершенствование процессов принятия управленческих и организационных решений в высших учебных заведениях. Повышение и рационализация эффективности процесса принятия решений. Более эффективное использование имеющихся ресурсов. Повышение эффективности образовательных программ.
Совершенствование работы приемной компании	Поиск и привлечение абитуриентов. Отбор наиболее квалифицированных абитуриентов, анализ имеющихся в открытых источниках сведений о школьниках-абитуриентах, поиск в списках победителей олимпиад и т.д.
Модернизация программно-технических средств обучения	Разработка системы параметров повышения эффективности веб-сервисов и их адаптации к пользователям (оптимальный размер сервера, распределение сетевого трафика и др.). Развитии и модернизация ИТ-инфраструктуры учреждений образования.
Оптимизация педагогического состава	Система оценки преподавателя, включая построение модели профиля преподавателя (тьютора). Формирование у профессорско-преподавательского состава компетенции по анализу больших данных.

В работе К.А. Баранникова, С.М. Лесина [9] говорится о возможных методах анализа цифрового следа обучения, при этом называются такие инструменты как регрессия, классификация, кластериза-

ции, ассоциативные правила, визуализация данных и др. Е.В. Ширинкина [10] называет такие методы анализа больших данных в образовании как интеллектуальный анализ причинно-следствен-

ных связей, статистические методы, поиск и выявление взаимосвязей при помощи моделей машинного обучения, модели прогнозирования и др. Обобщая сказанное и используя работы современных

авторов [11–12 и др.], можно выделить следующие направления применения больших данных в образовании (табл. 2).

### 3. Проблемы использования больших данных

Говоря о перспективах применения больших данных, нельзя не остановиться на проблемах, связанных с применением больших данных в образовании.

Е.М. Сергейчик в своей работе [13] отмечает, что «у больших данных есть и определенные недостатки, которые касаются прежде всего конфиденциальности и безопасности, поскольку возникает возможность манипуляции персональными данными, использования их для необоснованного надзора, нарушения неприкосновенности частной жизни». Там же говорится, что «создавая ситуацию неопределенности, сложности, нестабильности, неоднозначности, большие данные существенным образом изменяют наши представления о процессе познания и тем самым оказывают влияние на педагогические методы и принципы развития когнитивных способностей».

Разнообразные методологические и технические проблемы сбора и анализа больших данных в условиях цифровой экономики рассматриваются в работе [14]. При этом отмечается, что «анализ больших данных невозможен без технологических инструментов, развития технологических платформ для работы с базами данных, внедрение которых требует немалых средств и ресурсов, времени».

В статье [15] говорится о недостатках человеческого фактора и проблеме с кадрами при использовании больших данных, отмечается, что «использование больших данных требует определенной экспертизы и знаний в области статистики и анализа данных».

Ввиду этого можем заключить, что, несмотря на множество преимуществ, использование больших данных в образовании также сопряжено с рядом вызовов:

- защита личных (персональных) данных. Сбор и анализ больших объемов информации требует строгого соблюдения норм защиты персональных данных и их конфиденциальности. С этим связаны и этические дилеммы, вопросы о том, как этично использовать данные о студентах, что становится особенно актуальными в контексте принятия решений на основе аналитики;

- необходимость в новых методологиях и технологиях анализа больших данных, дальнейшее развитие таких моделей как MapReduce, NoSQL, Spark, платформы Hadoop и др., потребность в появлении новых высокоэффективных моделей распределенной обработки данных;

- потребность в существенной модернизации имеющихся в системе образования технических средств, в развитии ИТ-инфраструктуры на базе отечественных импортозамещающих информационных технологий;

- необходимость в квалифицированных кадрах. Эффективное использование аналитических инструментов требует наличия специалистов, способных интерпретировать данные и делать на их основе обоснованные выводы. На данный момент наблюдается нехватка специалистов в области анализа больших данных (Data Science), а также низкая потенциальная готовность педагогических работников к активному применению аналитики данных в ежедневной педагогической деятельности.

Большие данные имеют потенциал не только изменить подходы к обучению, но и повысить качество образования в целом. Они открывают возможности для персонали-

зации, анализа и оптимизации учебных процессов. Тем не менее, важно помнить о вызовах, связанных с защитой данных и этическими аспектами. Для успешной интеграции больших данных в образовательную практику необходимо не только высокотехнологическое техническое оснащение и создание соответствующей кадровой базы, но и формирование информационной культуры ответственности за использование персональных данных.

### 4. Пример анализа цифрового следа обучения с помощью моделей обработки больших данных Apache Spark

Apache Spark – это фреймворк с открытым исходным кодом для реализации распределенной обработки данных, входящий в систему проектов Hadoop. Он также поддерживает богатый набор высокоуровневых инструментов, включая Spark SQL для обработки SQL и структурированных данных, MLlib для машинного обучения, GraphX для обработки графов и Spark Streaming для потоковой обработки. Spark распределяет данные и вычисления по нескольким узлам в кластере, выполняя разные операции обработки параллельно в режиме реального времени.

Spark SQL поддерживает работу с различными источниками данных через интерфейс DataFrame. DataFrame можно использовать с помощью реляционных преобразований, а также для создания временного представления. Регистрация DataFrame в качестве временного представления позволяет выполнять SQL-запросы к его данным [16].

В ходе данного исследования был реализован тестовый пример анализа log-файла (журнала событий) электронного курса с помощью технологий обработки больших данных Spark SQL.

Лог-файл (log-file) электронного курса – это журнал событий, в который автоматически записываются все события, происходящие в компьютерной системе, включая все действия пользователей по работе с образовательным контентом.

С использованием фреймворка Apache Spark были созданы четыре DataFrame, представляющие фрагменты log-файла данных о взаимодействии студентов с элементами электронного курса. Структура логической модели разработанной базы данных DataFrame приводится на рис. 1.

Описание элементов электронного курса содержится в таблице *course*. Когда студент обращается к элементам курса, создается сеанс (сессия), которая заносится в таблицу *sessions*. Сессия может включать несколько обращений к элементам курса, их перечень содержится в таблице *reference*.

Таблица *course* содержит следующие сведения об элементах электронного курса:

- *id* – идентификатор элемента электронного курса;
- *name* – название элемента;
- *scores* – рейтинговые баллы активности за обращение к элементу;

Таблица *sessions* – данные о сеансах (сессиях) в системе:

- *id* – идентификатор сеанса;
- *session\_date* – дата сеанса;
- *student\_id* – идентификатор студента;

Таблица *reference* – данные об обращениях к элементам электронного курса:

- *course\_id* – идентификатор элемента курса;
- *session\_id* – идентификатор сеанса;
- *quantity* – количество обращений к элементу в ходе сеанса.

Таблица *student* – сведения о студентах:

- *id* – идентификатор студента;

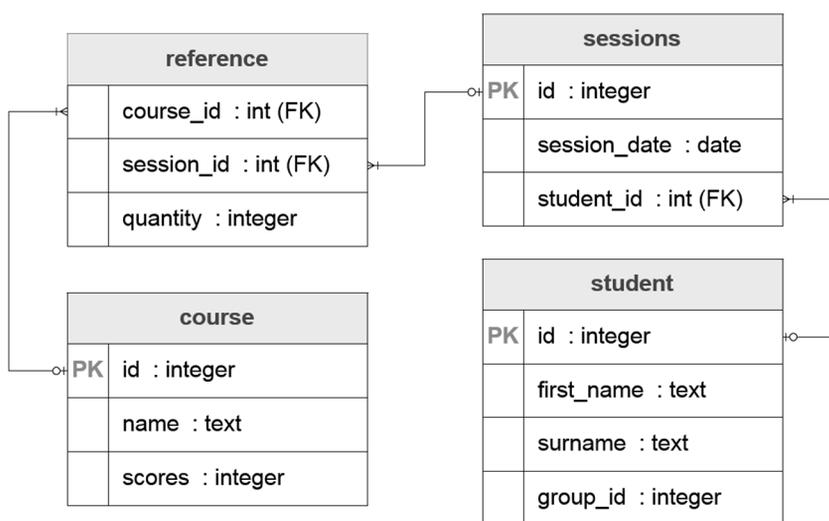


Рис. 1. Логическая схема разработанной на основе log-файла базы данных

Fig. 1. Logical diagram of the database developed on the basis of the log file

```
[ ] spark.sql("""SELECT id, name, session_id, quantity
FROM course
RIGHT JOIN reference
ON course.id == reference.course_id""").show(truncate=False)
```

id	name	session_id	quantity
1	Задание: Задания (Зачет/Незачет) Лабораторная работа. Создание	4	2
1	Задание: Задания (Зачет/Незачет) Лабораторная работа. Создание	5	1
3	Файл: Мет. указания Лабораторная работа. Запросы-действия. Итог	2	1
3	Файл: Мет. указания Лабораторная работа. Запросы-действия. Итог	6	3
5	Задание: Задания (Зачет/Незачет) Лабораторная работа. Разработк	7	5
5	Задание: Задания (Зачет/Незачет) Лабораторная работа. Разработк	8	5
4	Задание: Задания (Зачет/Незачет) Лабораторная работа. Кнопочные	1	7
4	Задание: Задания (Зачет/Незачет) Лабораторная работа. Кнопочные	3	1
2	Задание: Задания (Зачет/Незачет) Лабораторная работа. Программи	1	8

Рис. 2. Вид и результат Spark SQL запроса на поиск количества просмотров элементов курсов за каждую сессию (фрагмент)

Fig. 2. View and result of Spark SQL query to find the number of views of course elements for each session (fragment)

```
[ ] spark.sql("""SELECT id, name, session_id, quantity
FROM course
LEFT JOIN reference
ON course.id == reference.course_id
WHERE quantity is null
""").show(truncate=False)
```

id	name	session_id	quantity
6	Задание: Текущая аттестация (тестирование)	null	null
7	Файл: Лекция СУБД. Отчеты	null	null

Рис. 3. Вид и результат запроса на поиск элементов курса, не просмотренных ни одного раза

Fig. 3. View and result of the query to search for course elements that have not been viewed even once

- *first\_name* – имя;
- *surname* – фамилия;
- *group\_id* – идентификатор группы обучения.

Приведем примеры нескольких разработанных за-

просов Spark SQL для обработки логов электронного курса (рис. 2–4).

Также были созданы и реализованы Spark SQL запросы, позволяющие вычислять рей-

```
[ ] spark.sql("""SELECT session_date, student_id
                FROM sessions
                WHERE session_date == '2024-09-03'
                """).show(truncate=False)
```

session_date	student_id
2024-09-03	2014
2024-09-03	3254
2024-09-03	5784

**Рис. 4. Вид и результат запроса на поиск активности студентов в указанную дату**

**Fig. 4. View and result of the query to search for students' activity on the specified date**

тинг студентов с учетом рейтинговых баллов активности за обращение к элементам курса, находить элементы, внесшие наибольший вклад в рейтинг активности студентов и др.

Данные примеры показывают потенциальную возможность и практическую применимость технологий обработки больших данных к задачам анализа цифрового следа обучения.

### Заключение

В заключение следует отметить, что большие данные в образовании представляют собой инструмент, способный кардинально изменить подходы к обучению и оценке его эффективности. Анализ огромных массивов информации предоставляет уникальные возможности для выявления индивидуальных потребностей учащихся, оптимизации учебных процессов и разработки

адаптивных образовательных программ.

В ходе данной работы были рассмотрены ключевые аспекты использования больших данных в образовании, их источники в виде цифрового следа обучения, методы анализа и направления применения больших данных. Также был реализован тестовый пример анализа log-файла (журнала событий) электронного курса с помощью технологий обработки больших данных Spark SQL.

При этом были выделены следующие источники больших данных в образовании: электронная образовательная среда и электронная библиотека для обучения; сайт вуза; социальные сети и форумы; данные обратной связи, обращений и опросов; персональные данные, включая психометрические характеристики обучающихся; данные научных smart лабораторий; данные видеона-

блюдений и систем контроля управления доступом; данные о карьерном пути и успешности выпускников. Направления применения больших данных в образовании включают в себя следующие моменты: персонализация электронного обучения, выдача персонализированных рекомендаций; аналитика данных; оценка и обратная связь; прогнозирование успеха студентов; мониторинг качества образования; создание модели обучаемого; разработка учебных планов на основе запросов работодателей; разработка новых образовательных программ; появление новых моделей обучения; совершенствование процессов управления вузом; совершенствование работы приемной компании; модернизация программно-технических средств обучения; оптимизация педагогического состава.

Но нельзя также забывать, что использование больших данных в образовательной сфере также несет с собой определенные риски и вызовы, связанные с этическими аспектами, защитой личных данных и необходимостью кадровой модернизации сложившейся системы образования. Для успешной интеграции аналитики данных в образовательную практику необходимо развивать не только технические ресурсы, но и уровень цифровой безопасности и этики в использовании персональных данных.

### Литература

1. Baig M.I., Shuib L., Yadegaridehkordi E. Big data in education: a state of the art, limitations, and future research directions // *International Journal of Educational Technology in Higher Education*. 2020. Т. 17. № 1. С. 1–23. DOI: 10.1186/s41239-020-00223-0.
2. Lynch C. Big Data: How do your data grow? // *Nature*. 2008. Т. 455. № 7209. С. 28–29. DOI: 10.1038/455028a.
3. Шамсутдинова Т.М. Когнитивная модель траектории электронного обучения на основе цифрового следа // *Открытое образование*.

2020. Т. 24. № 2. С. 47–54. DOI: 10.21686/1818-4243-2020-2-47-54.

4. Огурцова Е.Ю., Фадеев Р.Н. Большие данные и цифровая аналитика в университетском образовании // *Ноосферные исследования*. 2021. № 4. С. 37–44. DOI: 10.46724/NOOS.2021.4.37-44.
5. Кондратенко А.Б., Кондратенко Б.А. Возможности применения больших данных в образовании в эпоху цифрового общества // *Вестник Калининградского филиала Санкт-Петербургского университета МВД России*. 2017. № 4(50). С. 112–115.

6. Аринушкина А.А., Тормосова А.К. Мониторинг обращений граждан в системе общего образования: электронное участие (e-participation) и «Большие данные» в образовании (Big Data in Education) // Человек и образование. 2019. № 4(61). С. 149–155.

7. Мамедова Г.А., Зейналова Л.А., Меликова Р.Т. Технологии больших данных в электронном образовании // Открытое образование. 2017. Т. 21. № 6. С. 41–48. DOI: 10.21686/1818-4243-2017-6-41-48.

8. Ермачкова Ю.В., Ливенец М.А., Селиверстова И.В. Использование больших данных при прогнозировании запроса рынка труда к системе образования // Труд и социальные отношения. 2021. Т. 32. № 6. С. 52–63.

9. Баранников К.А., Лесин С.М. Методология анализа больших данных в образовании (системно-методологический подход, основанный на анализе образовательных данных, поиска стратегии принятия управленческих и организационно-педагогических решений в образовании) // Народное образование. 2020. № 2(1479). С. 81–90.

10. Ширинкина Е.В. Data Driven как аналитика больших данных в образовании в условиях цифровизации // Качество и жизнь. 2022. № 2(34). С. 57–62.

11. Bai X., Zhang F., Li J., Guo T., Aziz A., Jin A., Xia F. Educational Big Data: Predictions, Applications and Challenges // Big Data Research. 2021. Т. 26. С. 100270. DOI: 10.1016/j.bdr.2021.100270.

12. Lee Y. Using self-organizing map and clustering to investigate problem-solving patterns in the massive open online course: an exploratory study // Journal of Educational Computing Research. 2019. Т. 57. № 2. С. 471–490. DOI: 10.1177/0735633117753364.

13. Сергейчик Е.М. Философско-антропологические аспекты проблемы «Больших данных» в образовании // Непрерывное образование. 2021. № 4(38). С. 4–10.

14. Лобан И.И. Проблемы сбора отчетности и образования больших данных в условиях цифровой экономики // Вестник Белорусской государственной сельскохозяйственной академии. 2022. № 2. С. 14–20.

15. Алиева М.В., Батчаева З.Б., Муцурова З.М., Исаева М.З. Большие данные и их применение в образовании // Журнал прикладных исследований. 2023. № 6. С. 140–146.

16. Spark SQL, DataFrames and Datasets Guide [Электрон. ресурс]. Режим доступа: <https://spark.apache.org/docs/3.5.3/sql-programming-guide.html>. (Дата обращения: 12.10.2024).

## References

1. Baig M.I., Shuib L., Yadegaridehkordi E. Big data in education: a state of the art, limitations, and future research directions. International Journal of Educational Technology in Higher Education. 2020; 17; 1; 44: 1-23. DOI: 10.1186/s41239-020-00223-0.

2. Lynch C. Big Data: How do your data grow? Nature. 2008; 455; 7209: 28-29. DOI: 10.1038/455028a.

3. Shamsutdinova T.M. Cognitive model of the e-learning trajectory based on the digital footprint. Otkrytoye obrazovaniye = Open Education. 2020; 24; 2: 47-54. DOI: 10.21686/1818-4243-2020-2-47-54. (In Russ.)

4. Ogurtsova Ye.Yu., Fadeyev R.N. Big Data and Digital Analytics in University Education. Noosfernyye issledovaniya = Noospheric Research. 2021; 4: 37-44. DOI: 10.46724/NOOS.2021.4.37-44. (In Russ.)

5. Kondratenko A.B., Kondratenko B.A. Possibilities of Using Big Data in Education in the Era of Digital Society. Vestnik Kaliningradskogo filiala Sankt-Peterburgskogo universiteta MVD Rossii = Bulletin of the Kaliningrad Branch of the St. Petersburg University of the Ministry of Internal Affairs of Russia. 2017; 4(50): 112-115. (In Russ.)

6. Arinushkina A.A., Tormosova A.K. Monitoring citizens' appeals in the general education system: electronic participation (e-participation) and «Big Data» in education. Chelovek i obrazovaniye = Man and education. 2019; 4(61): 149-155. (In Russ.)

7. Mamedova G.A., Zeynalova L.A., Melikova R.T. Big Data technologies in electronic education. Otkrytoye obrazovaniye = Open education. 2017; 21; 6: 41-48. DOI: 10.21686/1818-4243-2017-6-41-48. (In Russ.)

8. Yermachkova Yu.V., Livenets M.A., Seliverstova I.V. Using big data in forecasting labor market demand for the education system. Trud i sotsial'nyye otnosheniya = Labor and social relations. 2021; 32; 6: 52-63. (In Russ.)

9. Barannikov K.A., Lesin S.M. Methodology of Big Data Analysis in Education (a systemic and methodological approach based on the analysis of educational data, the search for a strategy for making managerial and organizational-pedagogical decisions in education). Narodnoye obrazovaniye = Public Education. 2020; 2(1479): 81-90. (In Russ.)

10. Shirinkina Ye.V. Data Driven as Big Data Analytics in Education in the Context of Digitalization. Kachestvo i zhizn' = Quality and Life. 2022; 2(34): 57-62. (In Russ.)

11. Bai X., Zhang F., Li J., Guo T., Aziz A., Jin A., Xia F. Educational Big Data: Predictions, Applications and Challenges. Big Data Research. 2021; 26: 100270. DOI: 10.1016/j.bdr.2021.100270.

12. Lee Y. Using self-organizing map and clustering to investigate problem-solving patterns in the massive open online course: an exploratory study. Journal of Educational Computing Research. 2019; 57; 2: 471-490. DOI: 10.1177/0735633117753364.

13. Sergeychik Ye.M. Philosophical and anthropological aspects of the problem of «Big Data» in education. *Nepreryvnoye obrazovaniye = Continuous education*. 2021; 4(38): 4-10. (In Russ.)

14. Loban I.I. Problems of reporting collection and big data generation in the digital economy. *Vestnik Belorusskoy gosudarstvennoy sel'skokhozyaystvennoy akademii = Bulletin of the Belarusian State Agricultural Academy*. 2022; 2: 14-20. (In Russ.)

15. Aliyeva M.V., Batchayeva Z.B., Mutsurova Z.M., Isayeva M.Z. Big data and their application in education. *Zhurnal prikladnykh issledovaniy = Journal of Applied Research*. 2023; 6: 140-146. (In Russ.)

16. Spark SQL, DataFrames and Datasets Guide [Internet]. Available from: <https://spark.apache.org/docs/3.5.3/sql-programming-guide.html>. (cited 12.10.2024).

#### **Сведения об авторе**

***Татьяна Михайловна Шамсутдинова***

*К.ф.-м.н, доцент кафедры цифровых технологий и прикладной информатики*

*Башкирский государственный аграрный университет, Уфа, Россия*

*Эл. почта: tsham@rambler.ru*

#### **Information about the author**

***Tatyana M. Shamsutdinova***

*Cand. Sci. (Physics and Mathematics), Associate Professor of the Department of Digital Technologies and Applied Informatics*

*Bashkir State Agrarian University, Ufa, Russia*

*E-mail: tsham@rambler.ru*